LOCAL SIMILARITY MEASURES FOR DEMONS-LIKE REGISTRATION ALGORITHMS

Antonio Tristán-Vega, Gonzalo Vegas-Sánchez-Ferrero and Santiago Aja-Fernández

Laboratory of Image Processing. Universidad de Valladolid. Spain

ABSTRACT

The main limitation of the original demons image registration algorithm is the need to ensure that the grey level of corresponding voxels remains the same in both the images to register. This difficulty may be overcome by the use of more robust similarity measures, whenever they may be expressed as the sum over the entire image of a local energy. We demonstrate the convenience of a local optimisation strategy of this energy instead of the optimisation of the global metric, and suggest a methodology for the use of more general similarity measures.

Index Terms— Image registration, biomedical imaging

1. INTRODUCTION

The demons algorithm was first introduced in [1] as an analogy with a thermodinamics paradox known as Maxwell's demons. It relies on the assumption that the grey level of corresponding voxels remains the same when the images are deformed, and tries to align the isocontours on both the images to register. This behaviour is implemented following the optical flow equations to compute a deformation field, which is further regularised by convolving it with a Gaussian kernel.

Demons-like algorithms have been succesfully applied to a wide variety of problems, such as US sequence tracking [2], US-MRI brain images corregistration [3], or even DT-MRI registration [4]. Moreover, the demons approach has been proved to be closely related to most of the registration algorithms based on dense deformation fields, such as optical flow [1], elastic body [5] and viscous fluid [6] models, or Iterative Closest Point [1], and a number of studies have been carried out to further improve its performance as well as to overcome its limitations.

In [7] it is demonstrated that the optical flow equation is equivalent to a Levenberg-Marquardt (LM) optimisation of the Mean Squared Difference (MSD) between the images to register; the dual formulation in [1], based either on the gradient of the fixed image or the gradient of the moving image, may be seen in such a way as a symmetric problem, depending on wether the deformation is applied to the fixed or the moving image. For the problem to be physically consistent, it is stated that both two approaches should be approximately equivalent. The main limitation of classical demons is the assumption that the grey levels of corresponding voxels do not vary from one image to the other. To overcome this drawback, in [8] a generalisation of the equation in [7] is obtained, so that the global similarity measure is defined as the sum of a local measure over the entire domain of the images, being the gradient of the moving image substituted by the gradient of the metric. When the local metric is the MSD, the equation reduces to the LM algorithm; for other metrics, the equivalency does not hold.

The estrategy followed in [8] consist in the use of a Local Correlation Coefficient (LCC), where the required local mean values and variances are computed as convolutions with Gaussian kernels. Then, the algorithm no longer needs the grey level to be the same in corresponding voxels, but instead it is enough that a linear relation between them holds. The computation of the gradient of the similarity measure is done in [8] in a global sense, which is, it is computed as the derivative of the global metric with respect to the displacement field. Given the complexity of the resulting equation, it is simplified by considering the Gaussian kernel narrow enough to neglect its effect. Although the results obtained with this approximation are systematically better than those for the original equation, a satisfactory explanation cannot be found in [8].

In this paper we demonstrate that the expression given in [8] is not only an approximation, but it corresponds instead to a parallel way in which the optical flow equation may be generalised. It is proved as well that this second methodology is a more adequate approach in cases of local, irregular deformations, and for global, smoother ones too. We also introduce a new local metric based on normalised MSD (NMSD), which shows better local behaviour than LCC while it still overcomes the limitation of MSD; its better performance when local derivatives are used suggests the validity of our work for more general problems, for instance with multimodal metrics.

In Section 2 we analyse the approach in [8], and we prove the equivalence of the approximate derivative given there with a different methodology based on local derivatives. In Section 3 we introduce the NMSD, compared to MSD and LCC. Section 4 is devoted to experimentally test the approaches here compared. Finally, in Section 5, we draw some final remarks and conclude.

This work was partially supported by grant numbers PI-041483 from the Fondo de Investigaciones Sanitarias, Spain; and TEC2007-67073/TCM from the Comisión Interministerial de Ciencia y Tecnología, Spain.

2. GENERALISATION OF THE LM APPROACH

The optical flow formulation given in [1], involving the gradient of both the fixed and the moving image, has been proved to be equivalent to an LM optimisation of the MSD [7]. Furthermore, considering only the gradient of the moving image, an equivalent equation is given in [8]:

$$\vec{v}[\mathbf{x}] = \frac{2E[\mathbf{x}]}{\|\nabla E[\mathbf{x}]\|^2 + 4\lambda E[\mathbf{x}]}\nabla E[\mathbf{x}], \qquad (1)$$

where $\vec{v}[\mathbf{x}]$ is the updating rule for the deformation field $\vec{u}[\mathbf{x}]$, say $\vec{v}[\mathbf{x}] = d\vec{u}[\mathbf{x}]/dt$, λ is the LM parameter, and $E[\mathbf{x}]$ is the locally defined similarity measure; in the case of the MSD, it is simply $E[\mathbf{x}] = (I[\mathbf{x}] - J[\mathbf{x}])^2$, where I stands for the fixed image and J for the moving image deformed with the current deformation field $\vec{u}$, so that MSD $= \sum_x E[\mathbf{x}]$. In [8] the LCC is used instead, and so the metric reads:

$$\text{LCC} = \sum_x E[\mathbf{x}] = \sum_x \frac{\overline{I[\mathbf{x}]J[\mathbf{x}]} - \bar{I}[\mathbf{x}]\bar{J}[\mathbf{x}]}{\sigma_I[\mathbf{x}]\sigma_J[\mathbf{x}]}, \qquad (2)$$

where $\bar{Y}$ denotes the mean value of Y. All mean values and statistics are computed as local spatial averages, *e. g.* $\bar{I}[\mathbf{x}] = \sum_s G_\eta[\mathbf{x}-\mathbf{s}]I[\mathbf{s}]$, with G_η a Gaussian kernel with isotropic variance η^2. The need to compute ∇E in eq. (1) is fulfilled in [8] by computing the derivatives of LCC with respect to the displacement field, and then we have, [8, eq. (2)]:

$$\begin{aligned}\frac{\partial \text{LCC}}{\partial \vec{u}[\mathbf{x}]} = &\left(I[\mathbf{x}] G_\eta * \frac{1}{\sigma_I\sigma_J}[\mathbf{x}] - J[\mathbf{x}] G_\eta * \frac{\overline{IJ} - \bar{I}\bar{J}}{\sigma_I\sigma_J^3}[\mathbf{x}] \right.\\ &\left. + G_\eta * \left(\frac{\left(\overline{IJ} - \bar{I}\bar{J}\right)\bar{J}}{\sigma_I\sigma_J^3} - \frac{\bar{I}}{\sigma_I\sigma_J} \right)[\mathbf{x}] \right) \nabla J[\mathbf{x}]. \qquad (3)\end{aligned}$$

On the other hand, we compute in what follows the derivatives of the local energy (similarity measure) $E[\mathbf{x}]$ with respect to the displacement field, and so, from eq. (2):

$$\begin{aligned}\frac{\partial E[\mathbf{x}]}{\partial \vec{u}[\mathbf{x}]} &= \frac{\partial}{\partial \vec{u}[\mathbf{x}]} \frac{\sum_s G_\eta[\mathbf{x}-\mathbf{s}]I[\mathbf{s}]J[\mathbf{s}] - \bar{I}[\mathbf{x}]\bar{J}[\mathbf{x}]}{\sigma_I[\mathbf{x}]\sigma_J[\mathbf{x}]}\\
&= \frac{G_\eta[\mathbf{0}]I[\mathbf{x}]}{\sigma_I[\mathbf{x}]\sigma_J[\mathbf{x}]}\nabla J[\mathbf{x}] - \frac{\bar{I}[\mathbf{x}]}{\sigma_I[\mathbf{x}]\sigma_J[\mathbf{x}]}\frac{\partial \bar{J}[\mathbf{x}]}{\partial \vec{u}[\mathbf{x}]}\\
&\quad - \frac{1}{2}\frac{\overline{I[\mathbf{x}]J[\mathbf{x}]} - \bar{I}[\mathbf{x}]\bar{J}[\mathbf{x}]}{\sigma_I[\mathbf{x}]\sigma_J^3[\mathbf{x}]}\frac{\partial \sigma_J^2[\mathbf{x}]}{\partial \vec{u}[\mathbf{x}]}\\
&= \frac{G_\eta[\mathbf{0}]I[\mathbf{x}]}{\sigma_I[\mathbf{x}]\sigma_J[\mathbf{x}]}\nabla J[\mathbf{x}] - \frac{\bar{I}[\mathbf{x}]G_\eta[\mathbf{0}]}{\sigma_I[\mathbf{x}]\sigma_J[\mathbf{x}]}\nabla J[\mathbf{x}]\\
&\quad - \frac{E[\mathbf{x}]}{2\sigma_J^2[\mathbf{x}]}\frac{\partial}{\partial \vec{u}[\mathbf{x}]}\left(\sum_s G_\eta[\mathbf{x}-\mathbf{s}]J^2[\mathbf{s}] - \bar{J}^2[\mathbf{s}]\right)\\
&= \frac{G_\eta[\mathbf{0}]I[\mathbf{x}]}{\sigma_I[\mathbf{x}]\sigma_J[\mathbf{x}]}\nabla J[\mathbf{x}] - \frac{\bar{I}[\mathbf{x}]G_\eta[\mathbf{0}]}{\sigma_I[\mathbf{x}]\sigma_J[\mathbf{x}]}\nabla J[\mathbf{x}]\\
&\quad - \frac{E[\mathbf{x}]}{2\sigma_J^2[\mathbf{x}]}2G_\eta[\mathbf{0}]J[\mathbf{x}]\nabla J[\mathbf{x}]\\
&\quad + \frac{E[\mathbf{x}]}{2\sigma_J^2[\mathbf{x}]}2G_\eta[\mathbf{0}]\bar{J}[\mathbf{x}]\frac{\partial}{\partial \vec{u}[\mathbf{x}]}\sum_s G_\eta[\mathbf{x}-\mathbf{s}]J[\mathbf{s}]\\
&= \frac{G_\eta[\mathbf{0}]}{\sigma_I[\mathbf{x}]\sigma_J[\mathbf{x}]}\left(I[\mathbf{x}] - \bar{I}[\mathbf{x}]\right)\nabla J[\mathbf{x}]\\
&\quad - \frac{E[\mathbf{x}]}{\sigma_J^2[\mathbf{x}]}G_\eta[\mathbf{0}]\left(J[\mathbf{x}] - \bar{J}[\mathbf{x}]\right)\nabla J[\mathbf{x}], \qquad (4)\end{aligned}$$

given that $\partial J[\mathbf{s}]/\partial \vec{u}[\mathbf{x}] = \delta[\mathbf{x}-\mathbf{s}]\nabla J[\mathbf{x}]$. Bearing in mind the definition of $E[\mathbf{x}]$ in eq. (2), we get:

$$\begin{aligned}\frac{\partial E[\mathbf{x}]}{\partial \vec{u}[\mathbf{x}]} = &\frac{G_\eta[\mathbf{0}]}{\sigma_I[\mathbf{x}]\sigma_J[\mathbf{x}]}\left(\left(I[\mathbf{x}] - \bar{I}[\mathbf{x}]\right)\right.\\
&\left. - \left(J[\mathbf{x}] - \bar{J}[\mathbf{x}]\right)\frac{\overline{I[\mathbf{x}]J[\mathbf{x}]} - \bar{I}[\mathbf{x}]\bar{J}[\mathbf{x}]}{\sigma_J^2[\mathbf{x}]}\right)\nabla J[\mathbf{x}], \qquad (5)\end{aligned}$$

which is identical to [8, eq. (3)] except for the term $G_\eta[\mathbf{0}]$. The meaning of this result is that in fact eq. (5) is not only an approximation to the derivative of the global LCC similarity measure given in eq. (3) for narrow Gaussian kernels, as claimed in [8]. It corresponds instead to the derivative of the local energy function, and therefore it drives to a different way to generalise eq. (1). In the case of the MSD, it is trivial to show that both two methodologies gives the same approach (the LM equation), since the local energy for MSD depends only on the voxel of interest x. For any other local energy depending on the vicinity of x, this will not be the case.

As a final remark, let us make explicit the time dependencies; the local energy at each image voxel and time step is a function of the value of both images to register at this particular place and instant, i. e. $E[\mathbf{x}, t] = f(I[\mathbf{x}], J[\mathbf{x}, \vec{u}[\mathbf{x}, t]]) = g[\vec{u}[\mathbf{x}, t]]$. The time dependency of the local energy relies on the fact that the moving image J evolves with time following the deformation given by $\vec{u}$, which in turn is updated at each time step with the correction term $\vec{v}$. Then, we have:

$$\begin{aligned}\frac{dE[\mathbf{x}]}{dt} &= \frac{dg[\vec{u}[\mathbf{x}, t]]}{dt} = \nabla^t g[\vec{u}[\mathbf{x}, t]]\frac{d\vec{u}[\mathbf{x}, t]}{dt}\\
&= \nabla^t E[\mathbf{x}]\vec{v}[\mathbf{x}] = \frac{2\|\nabla E[\mathbf{x}]\|^2 E[\mathbf{x}]}{\|\nabla E[\mathbf{x}]\|^2 + 4\lambda E[\mathbf{x}]}. \qquad (6)\end{aligned}$$

For those image locations where $\|\nabla E[\mathbf{x}]\|^2 \gg 4\lambda E[\mathbf{x}]$ (note that the LM regularisation parameter λ should be small), eq. (6) reduces to $dE[\mathbf{x}]/dt = C \cdot E[\mathbf{x}]$, whose solution is an increasing exponential function that stops growing when the optimum is reached and $\|\nabla E[\mathbf{x}]\|^2 \gg 4\lambda E[\mathbf{x}]$ does not hold. Although eq. (1) is applied in [8] in a heuristic way, it yields a very fast rate of convergence with an adequate similarity measure. Note that this is only valid in case the local derivative is used, which is a clear reason that favours the use of the methodology presented in our paper, together with the simpler mathematical expression of $\vec{v}[\mathbf{x}]$. On the other hand, note that LCC must be maximised, as opposed to MSD,

for which a minus sign in eq. (1) is needed and we have $dE[\mathbf{x}]/dt = -C \cdot E[\mathbf{x}]$, with an exponential decrease; it explains why the same expression of MSD stands for LCC.

3. NORMALISED MEAN SQUARED DIFFERENCE

The LCC is able to cope with varying grey levels, which is a clear advantage compared to the MSD. However, as it was showed in the previous section, the computation of statistics carries out a worsening of the local behaviour. For two random variables I and J, it is trivial to show that the normalised correlation relates to the expected value of the MSD between normalised versions of the original I and J, which is:

$$\mathrm{CC}(I,J) = \frac{\overline{IJ} - \bar{I}\bar{J}}{\sigma_I \sigma_J} = 1 - \frac{1}{2}\overline{\left(\frac{I-\bar{I}}{\sigma_I} - \frac{J-\bar{J}}{\sigma_J}\right)^2}. \quad (7)$$

The normalisation of I and J yields transformed random variables $\tilde{I}$ and $\tilde{J}$ with zero mean and unit variance, which makes this similarity measure robust against linear changes in the grey levels. In this context, the expected values must be seen as local spatial averages which are responsible that the LCC is not perfectly local. Such averages are strictly necessary to compute local statistics of the images, but we can avoid the local smoothing due to the computation of the average of the normalised MSD. This way, we introduce the normalised MSD, including now the spatial dependencies, as:

$$\mathrm{NMSD} = \sum_x \left(\frac{I[\mathbf{x}] - \bar{I}[\mathbf{x}]}{\sigma_I[\mathbf{x}]} - \frac{J[\mathbf{x}] - \bar{J}[\mathbf{x}]}{\sigma_J[\mathbf{x}]}\right)^2. \quad (8)$$

Like in eq. (4) to (5), it can be easily proved that the derivative of the local energy for this metric follows:

$$\frac{\partial E}{\partial \vec{u}} = -\frac{2\sqrt{E}}{\sigma_J}\left(1 - G_\eta[\mathbf{0}] - G_\eta[\mathbf{0}]\left(\frac{J-\bar{J}}{\sigma_J}\right)^2\right)\nabla J, \quad (9)$$

where we have obviated any dependency with $\mathbf{x}$. The derivative of the global measure, $\partial \mathrm{NMSD}/\partial \vec{u}[\mathbf{x}]$, is slightly more complicated, involving convolution terms like in [8], and is not depicted here, but its computation is straightforward as well. Due to the normalisation of the images, NMSD is able to deal with linear changes in the grey level of the images, but it is slightly more local than LCC, being a compromise between LCC and the classical MSD.

4. RESULTS

To test the proposed methodology, we have used the public BrainWeb database[1] [9]. It comprises 20 3D anatomical models of a normal human brain, together with T1 MRI realistic simulations for 19 up of the 20 models. Three sets of experiments have been designed. To test the accuracy of the algorithm in cases of irregular or very localised deformations, we first use one of the T1 volumes (case 04) as target -fixed- image, and then we register each of the remaining 18 volumes using them as template -moving- images (**E1**). As in [8], we design a second experiment **E2** in which a linear bias is introduced in the target image (case 04), in order to test the robustness against variations of the grey level, and then we repeat the previous methodology. The bias is introduced by adding a correction ranging from 0 in the top-left corner of the first slice to 150 in the bottom-right of the last one, and then normalising the intensities in the range $[0, 255]$. Finally, to test the behaviour against smoother deformations, we have designed a third experiment **E3** in which we only use case 04: we place 27 landmarks uniformly distributed over the $181 \times 217 \times 181$ volume, and generate random displacements that are later on interpolated with a Thin Plate Spline (TPS), so that the achieved deformation has mean values near 5 voxels (5 mm.) and maximum values near 12 voxels (12 mm.). We generate 18 random deformations of case 04 with no bias.

To measure the performances, we can use the **euclidean error** $\mathcal{E}$ committed only in the third experiment. The **MSD** with respect to the unbiased T1 target may be computed in all cases. Besides, taking advantage of the golden standard segmentation provided by the anatomical models, we may use the Jaccard overlap measure **S** to account for the misalignment of corresponding tissues on both of the images:

$$\mathrm{S} = \frac{1}{M}\sum_{x \in Z} \Omega\left(M_I[\mathbf{x}], M_J[\mathbf{x}]\right), \quad (10)$$

where M_I, M_J are the anatomical models giving the kind of tissue at location $\mathbf{x}$, and Ω is an indicator, whose value is 0 if $M_I[\mathbf{x}] = M_J[\mathbf{x}]$ and 1 otherwise. The sum is done over a strip Z of 9 voxels width, containing all the interfaces between the tissues, a total of M voxels. We do so to obtain a more distinctive measure that accounts for the overlap only near the borders of the tissues, since otherwise the overlap of large, homogeneous tissues could bias the result.

We have tested 5 algorithms: one for the MSD measure, two for LCC with local -LCC_l- and global -LCC_g- computed derivatives and, respectively, NMSD_l and NMSD_g. In all cases, we empirically set $\eta = 4.1$ mm., and used a multirresolution scheme with 3 levels and 20 iterations per level. The deformation field was regularised at each step with a Gaussian kernel with $\sigma = 0.80$ mm. The results, averaged over the 18 experiments for each case, are shown in Table 1.

First of all, note that all performance measures agree in determine which metric is better than one another in all cases. Without bias, they may be ordered, from the best to the worst, as: MSD, NMSD_l, LCC_l, NMSD_g, and LCC_g. When we apply the bias, the MSD completely fails to register the images, but the LCC and NMSD show a performance very close to the unbiased case. Moreover, the previous order holds independently on the kind of deformation: the more local the

[1] http://www.bic.mni.mcgill.ca/brainweb/

E1	MSD	LCC_l	LCC_g	$NMSD_l$	$NMSD_g$
MSD	**16.4**	21.2	28.9	19.8	27.6
S	**0.345**	0.359	0.417	0.356	0.406

E2	MSD	LCC_l	LCC_g	$NMSD_l$	$NMSD_g$
MSD	*37.6*	21.2	28.9	**19.8**	27.6
S	*0.601*	0.359	0.416	**0.356**	0.405

E3	MSD	LCC_l	LCC_g	$NMSD_l$	$NMSD_g$
MSD	**8.3**	10.2	17.1	10.0	15.2
S	**0.185**	0.214	0.319	0.204	0.296
$\mathcal{E}$ (mm.)	**0.619**	1.40	2.37	1.13	2.12

Table 1. MSD and S for the inter-patient experiments, both without (E1) and with bias (E2), and for the synthetic deformations (E3). In E3, $\mathcal{E}$ can also be measured.

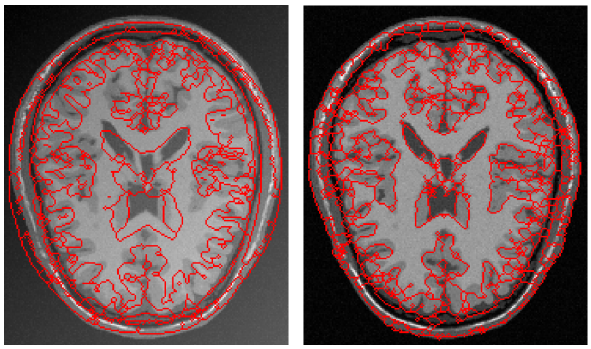

Fig. 1. An example of the performance of the algorithm with LCC_l and for E2. **Left**: axial view of the fixed image with bias, with the contours of the moving image superimposed. **Right**: the same axial view without the bias, with the contours of the registered image superimposed.

similarity measure and the computed derivative, the better the performance, both for smooth, global deformations and for very localised ones. The heavy difference between local and global derivatives, consistently with eq. (6), demonstrate the usefulness of our approach, and completely justifies the previous results in [8], which have been proved to generalise for other metrics than LCC and for a 3D case as well. On the other hand, the difference between NMSD and LCC is not so important; the computation of local statistics worsens the locality of NMSD, whose behaviour is far from the MSD. An example of the performance for E2 is shown in Fig. 1.

5. CONCLUSION

We have presented a methodology for the generalisation of the optical flow equation with non-MSD metrics. At the sight of the results, we may conclude that the use of local similarity measures combined with local optimisation rules drives not only to more simple equations (and therefore to more computationally efficient algorithms), but it also increases the accuracy of the registration, which is, the more the locality, the better the performance. On the other hand, it is necessary to achieve a trade-off between the locality of the metric and its robustness against changes in the grey levels. In this sense, we intend to use this result to extend the demons algorithm to the multimodal case; the main difficulty in this task is the need to define an adequate local similarity measure suitable for multimodal registration, since a functional relation between grey levels more general than a linear correspondence (like with LCC and NMSD) requires more robust metrics.

6. REFERENCES

[1] J.-P. Thirion, "Image matching as a diffusion process: an analogy with Maxwell's demons," *Med. Im. Anal.*, vol. 2, no. 3, pp. 243–260, Sept. 1998.

[2] X. Pennec, P. Cachier, and N. Ayache, "Tracking brain deformations in time sequences of 3D US images," *Patt. Recogn. Lett.*, vol. 24, no. 4–5, pp. 801–813, 2003.

[3] A. Guimond, A. Roche, N. Ayache, and J. Meunier, "Three-dimensional multimodal brain warping using the demons algorithm and adaptive intensity corrections," *IEEE Trans. on Med. Imag.*, vol. 20, no. 1, pp. 58–69, 2001.

[4] A. Guimond, C.R.G. Guttmann, S.K. Warfield, and C.-F. Westin, "Deformable registration of DT-MRI data based on transformation invariant tensor characteristics," in *IEEE Int. Sym. on Biomed. Imag.*, 2002, pp. 761–764.

[5] G.E. Christensen, M.I. Miller, and M.W. Vannier, "A 3-D deformable magnetic resonance textbook based on elasticity," in *AAAI Spring Symposium*, Palo Alto, CA, March 1994, pp. 153–156.

[6] M. Bro-Nielsen and C. Gramkow, "Fast fluid registration of medical images," in *Procs. of VBC '96*, London, UK, 1996, pp. 267–276, Springer-Verlag.

[7] X. Pennec, P. Cachier, and N. Ayache, "Understanding the "demon's algorithm": 3D non-rigid registration by gradient descent," in *Procs. of MICCAI '99*, London, UK, 1999, pp. 597–605, Springer-Verlag.

[8] P. Cachier and X. Pennec, "3D non-rigid registration by gradient descent on a Gaussian-windowed similarity measure using convolutions," in *Procs. of MMBIA '00*, Hilton Head Island, South Carolina, USA, June 2000, pp. 182–189, IEEE Computer society.

[9] D.L. Collins, A.P. Zijdenbos, V. Kollokian, J.G. Sled, N.J. Kabani, C.J. Holmes, and A.C. Evans, "Design and construction of a realistic digital brain phantom," *IEEE Trans. on Med. Imag.*, vol. 17, no. 3, pp. 463–468, 1998.